Operationalising Relative Causal Knowledge: Backbone Identifiability from Private Reports

Fabrizio Russo[1], Mark Somers[2]

[1]Imperial College London, London, United Kingdom
[2]Fifty One Degrees Ltd, London, United Kingdom

Abstract

The *Relativity of Causal Knowledge* (RCK) explains how a network of agents with different structural causal models can exchange causal knowledge through a shared interventionally consistent abstraction, or backbone. We ask the prior identification question that this transport mechanism presupposes: when is that backbone determined by the agents' private causal knowledge? In the basic two-agent common-effect case, two private causes influence one shared outcome and each agent identifies only the single-cause causal marginal relevant to its own perspective. We show that, under standard compatibility, non-degeneracy, and local overlap assumptions, those local causal marginals do not identify a unique backbone. Infinitely many joint intervention kernels can induce exactly the same private reports while disagreeing on joint interventions. We then give a conditional recovery result. Additive separability removes the hidden interaction degree of freedom, but observational residual summaries remain insufficient. Identification becomes possible when agents communicate causally identified response functions. An education value-added example illustrates why this is first a communication problem, and only then a policy-composition problem.

Keywords

causal knowledge, causal abstraction, identifiability, multi-agent causal inference, causal transport

1. Introduction

Causal knowledge is useful partly because it travels. A randomised trial, a quasi-experimental design, or a credible structural model does not merely answer one isolated question; it can become evidence for later decisions, related domains, or other researchers' models. Standard structural causal models (SCMs) provide the language of interventions and counterfactuals within one model [1, 2]. The *Relativity of Causal Knowledge* (RCK) framework [3] pushes this idea into a distributed setting: different agents may hold different local SCMs for the same world, and causal knowledge can be transported between them if their local perspectives admit a shared interventionally consistent abstraction, called a *backbone*.

That conditional is powerful, but it leaves an operational question open. In applications, the backbone is usually not available to the agents. It must be inferred from the knowledge they privately possess. If that inference is underdetermined, two agents can each report a valid causal finding and still fail to identify the shared abstraction through which those findings can be shared accurately. RCK is mathematically defined *given* a backbone, but the choice of backbone becomes an identification problem.

Figure 1 separates the transport operation from the identification question studied here. Agent ρ and Agent σ are nodes in a network that can probe the same system, by intervention or just observation, from different local perspectives, producing causal-knowledge values $\chi^\rho \in CK(M^\rho)$ and $\chi^\sigma \in CK(M^\sigma)$. $CK(M)$ denotes the causal-knowledge object associated with a SCM M, which lives on the edge stalks of the network. If a common edge value $\chi^\tau \in CK(M^\tau)$ is available, RCK can restrict local knowledge to the backbone τ through *restriction maps* ($\alpha_\chi^{\rho\triangleleft\tau}$ and $\alpha_\chi^{\sigma\triangleleft\tau}$), align it there, and extend it into another perspective through *extension maps* ($\beta_\chi^{\sigma\triangleleft\tau}$). Our question is the prior one: do the agents' private reports determine the backbone value χ^τ that this restriction–alignment–extension pipeline needs? The following example instantiates the diagram with two local findings about the same outcome.

KR / FLoC 2026 Workshop - JoeFest: A Workshop in Honor of Joseph Y. Halpern, July 19, 2026, Lisbon, Portugal
fabrizio@imperial.ac.uk (F. Russo); mark.somers@outlook.com (M. Somers)

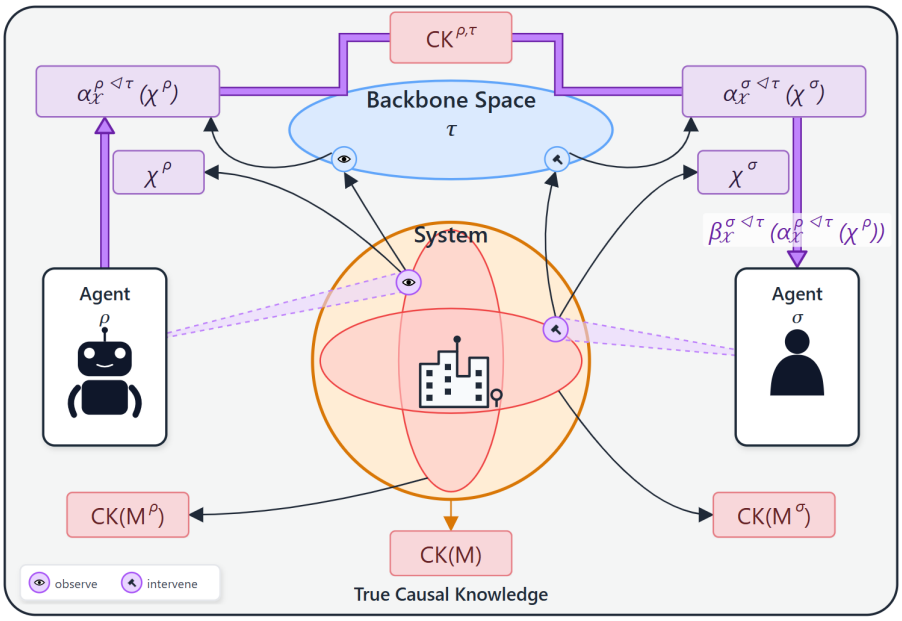
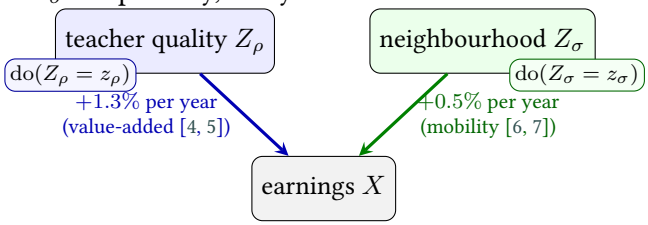


Figure 1: RCK transport assumes a shared edge value in the backbone space τ. The purple route shows local knowledge being restricted to the edge, aligned there, and extended into another agent's perspective; this paper asks when private reports identify that edge value in the first place.

Example 1 (Education value-added)**.** Consider a policy-maker asking whether better teachers can compensate for childhood neighbourhood disadvantage, e.g. in the context of the pupil premium funds in UK.[1] Adult earnings X are a natural shared outcome: the teacher value-added literature estimates long-run earnings effects of teacher quality [4, 5], while the neighbourhood-mobility literature estimates long-run earnings effects of childhood environment [6, 7]. Let Agent ρ study teacher quality Z_ρ, and let Agent σ study neighbourhood environment Z_σ. Empirically, one year with a teacher one standard deviation higher in value-added is associated with roughly 1.3% higher adult earnings, while one childhood year in a one-standard-deviation better county is associated with roughly 0.5% higher adult earnings [5, 7]. Both local findings can be correct. In the notation of Figure 1, they are local pieces of χ^ρ and χ^σ: the teacher study reports how X changes under interventions on Z_ρ, and the neighbourhood study reports the analogous response for Z_σ. What they do not determine is the edge value χ^τ in the backbone. The RCK consequence is that communication itself becomes ambiguous. To use Agent ρ's teacher finding inside Agent σ's perspective, the agents must know which backbone value it is being extended through; but several candidate χ^τ values can agree with both local findings while disagreeing on a joint intervention $\mathrm{do}(Z_\rho = z_\rho, Z_\sigma = z_\sigma)$. Thus Agent ρ's causal knowledge has no unique interventionally meaningful translation for Agent σ, and any compensatory policy calculation inherits that arbitrary choice.



Our contribution is to make this missing communication step precise. First, we formulate backbone recovery as an edge-level identification problem in the language of RCK. Second, for common-effect backbones, we prove a kernel-level non-identifiability theorem: under explicit non-degeneracy and local minorisation conditions, local causal marginals can leave infinitely many causally distinct backbones compatible with the same private reports, so the agents do not know which edge value they are aligning on. Third, we show how the obstruction can be removed under additive separability, provided that agents communicate causally identified response functions rather than merely observational summaries.

[1] https://www.gov.uk/government/publications/pupil-premium/pupil-premium

Position in the Literature. The closest point of departure is RCK itself [3]. RCK studies when local SCMs admit a common interventionally consistent abstraction and how causal knowledge can be transported across the resulting *network sheaf and cosheaf of causal knowledge.* Related causal-abstraction work studies exact transformations, semantic embedding, and compositionality across levels of description [8, 9, 10, 11, 12]. Our question is earlier. We do not ask whether transport is well behaved once an abstraction is available, nor whether an abstraction can be learned between specified models.

The problem is also distinct from aggregating expert causal judgements outside the abstraction setting. Bradley et al. [13] study aggregation rules for causal-network judgements and associated probabilities over a common variable set. Alrajeh et al. [14] instead take experts' causal models as inputs and define compatibility and dominance conditions under which those models can be merged, while Friedenberg and Halpern [15] extend this programme to experts with different focus areas using a *can-explain* relation. Our inputs and target are different. We do not assume access to full expert models, structural equations, or focus sets, and we are not trying to select a collective graph or merged SCM. We ask whether local causal reports determine the RCK edge object through which transport maps would apply; the goal is the backbone identifiability rather than graph or model aggregation.

Finally, transportability, data fusion, and federated causal inference study how evidence from different environments or sources can be combined under explicit assumptions [16, 17, 18, 19]. Causal marginal and latent-variable compatibility work similarly warns that partial causal views need not determine a unique joint source [20, 21]. We share that intuition, but the target differs: the unidentified object is not a pooled treatment effect, a merged graph, or a full joint SCM. It is the edge in the network sheaf and cosheaf through which RCK would allow sharing abstracted causal knowledge.

2. Backbone Identification from Colliding Private Reports

Recall the two-agent scenario of Example 1. Agent ρ has a private cause Z_ρ, Agent σ has a private cause Z_σ, and both care about the same outcome X. This creates a common-effect, collider configuration $Z_\rho \to X \leftarrow Z_\sigma$: the backbone answers to joint interventions on both private causes. In this setting a mixed response coordinate can be hidden from both one-cause reports. In RCK, the backbone is the shared edge object on which local causal knowledge must agree before it can be transported. In this two-agent setting, that edge value is represented concretely by the interventionally consistent kernel

$$Q(H \mid z_\rho, z_\sigma) = P(X \in H \mid \mathrm{do}(Z_\rho = z_\rho, Z_\sigma = z_\sigma)), \tag{1}$$

for measurable outcome events H. This joint-intervention kernel represents the edge value χ^τ. With χ^τ fixed, restriction and extension maps have a common object through which to transport abstracted causal knowledge. If the other private cause is not part of the agent's local design, a candidate backbone Q projects to single-cause reports by averaging over that hidden private cause:

$$P_\rho(\cdot \mid z_\rho) := \int Q(\cdot \mid z_\rho, z_\sigma)\, P_{Z_\sigma}(dz_\sigma), \qquad P_\sigma(\cdot \mid z_\sigma) := \int Q(\cdot \mid z_\rho, z_\sigma)\, P_{Z_\rho}(dz_\rho). \tag{2}$$

These *local report operators* describe how private information hides the other agent's knowledge. The identification problem fixes the reported kernels (P_ρ, P_σ) and asks whether they determine a unique, interventionally consistent Q. Thus Eq. (2) should be read as a report-projection constraint: a candidate backbone $\bar{Q}$ is compatible if

$$\int \bar{Q}(\cdot \mid z_\rho, z_\sigma)\, P_{Z_\sigma}(dz_\sigma) = P_\rho(\cdot \mid z_\rho), \qquad \int \bar{Q}(\cdot \mid z_\rho, z_\sigma)\, P_{Z_\rho}(dz_\rho) = P_\sigma(\cdot \mid z_\sigma).$$

Backbone identification requires all compatible candidates to agree on the intervention queries of interest. If two compatible kernels agree with both private reports but disagree on a joint intervention, then the edge value χ^τ needed for RCK transport is not identified.

Generic Non-Identifiability. We prove that such compatible kernels are generically not unique under weak assumptions: First, the two reports must be compatible with at least one candidate backbone Q_0; Second, each private-cause space must contain a nontrivial centred direction: bounded nonzero

functions $u(Z_\rho)$ and $v(Z_\sigma)$ with mean zero under the private-cause laws P_{Z_ρ} and P_{Z_σ}; Third, Q_0 must satisfy a local overlap requirement, formulated as a minorisation or small-set condition [22, Sec. 5.1], so that a small signed perturbation can be added without leaving the space of probability kernels.

Theorem 1 (Backbone non-identifiability under private knowledge)**.** *Assume:*

1. *there exists at least one compatible backbone kernel Q_0;*
2. *there are measurable sets A_ρ and A_σ with positive private-cause probability, and bounded nonzero mean-zero functions $u(Z_\rho)$ and $v(Z_\sigma)$ supported on A_ρ and A_σ, respectively;*
3. *for some probability measure ν with $0 < \nu(B) < 1$ for some measurable B, and some $\varepsilon > 0$, $Q_0(H \mid z_\rho, z_\sigma) \geq \varepsilon\, \nu(H)$ for every measurable event H and all $(z_\rho, z_\sigma) \in A_\rho \times A_\sigma$.*

Then there is a radius $\delta > 0$ and a one-parameter family of distinct kernels $\{Q_t : t \in (-\delta, \delta)\}$ such that every Q_t induces the same private reports P_ρ and P_σ, but for $t \neq s$ the kernels Q_t and Q_s disagree on the joint interventional distribution for some intervention pair (z_ρ, z_σ).

Proof sketch. Start from one compatible kernel Q_0. The proof has three steps. First, choose a nontrivial signed perturbation S on the outcome space with total mass zero, so adding it can change some events without breaking normalisation. Second, multiply this perturbation by centred functions $u(z_\rho)$ and $v(z_\sigma)$ on the two private-cause spaces. This makes the perturbation disappear whenever either agent averages over the other agent's private cause. Third, use the local minorisation condition to keep the perturbed laws non-negative for sufficiently small amplitudes. Concretely, define

$$Q_t(\cdot \mid z_\rho, z_\sigma) = Q_0(\cdot \mid z_\rho, z_\sigma) + t\, u(z_\rho) v(z_\sigma) S(\cdot). \tag{3}$$

The product $u(z_\rho)v(z_\sigma)$ is the hidden interaction direction: it is invisible to both one-cause reports because each report integrates out one centred factor, e.g. $\int u(z_\rho)v(z_\sigma)P_{Z_\sigma}(dz_\sigma) = 0$ and symmetrically for Z_ρ. But the same product is visible before marginalisation, and because u, v, and S are nontrivial, two different amplitudes $t \neq s$ disagree for some joint intervention pair. Thus the same private reports are compatible with infinitely many causally distinct backbone candidates. □

In RCK terms, the failure occurs at the edge stalk level. The agents' private reports can agree with many candidate edge values (represented by kernels Q_t), so there is no unique common χ^τ through which restriction, alignment, and extension should proceed. The following linear witness makes the communication failure concrete.

Example 2 (Why naive communication fails in the education case)**.** A minimal witness is

$$X = \alpha Z_\rho + \beta Z_\sigma + \gamma Z_\rho Z_\sigma + \varepsilon, \qquad \mathbb{E}[Z_\rho] = \mathbb{E}[Z_\sigma] = 0.$$

The single-cause intervention responses identify α and β: $\mathbb{E}[X \mid \mathrm{do}(Z_\rho = z_\rho)] = \alpha z_\rho, \mathbb{E}[X \mid \mathrm{do}(Z_\sigma = z_\sigma)] = \beta z_\sigma$, respectively. They do not identify γ. Thus the same teacher and neighbourhood reports are compatible with compensatory effects ($\gamma < 0$), additive effects ($\gamma = 0$), or complementary effects ($\gamma > 0$). The agents can exchange truthful local summaries, but those summaries do not specify a unique backbone value and therefore do not support unambiguous RCK communication. In policy terms, this leaves open whether better teachers help most in disadvantaged neighbourhoods, help equally everywhere, or are amplified by neighbourhood advantage.

A Conditional Recovery Result. The perturbation in Eq. (3) is invisible locally because it is a mixed term in the two private causes. A natural way to rule out that ambiguity is additive separability:

$$X = f_\rho(Z_\rho) + f_\sigma(Z_\sigma) + \varepsilon. \tag{4}$$

This assumption should not be read as a free modelling convenience. In the education example, it is motivated by the way the two literatures are usually parameterised and by the institutional separation between classroom instruction and neighbourhood or family environment [23, 24, 25]. It is also empirically revisable; interactions can exist and should be tested where joint data are available [26]. The role of separability here is precise: it removes the hidden interaction coordinate that drives Theorem 1.

Proposition 2 (Additive separability removes the mixed response)**.** *If the shared outcome X satisfies Eq. (4) and $\mathbb{E}[\varepsilon] < \infty$, then the joint interventional mean response is*

$$m(z_\rho, z_\sigma) := \mathbb{E}[X \mid \mathrm{do}(Z_\rho = z_\rho, Z_\sigma = z_\sigma)] = f_\rho(z_\rho) + f_\sigma(z_\sigma) + \mathbb{E}[\varepsilon].$$

In particular, the response contains no mixed term in (z_ρ, z_σ).

Separability is the structural half of recovery, but it does not imply identifiability. Even with the structural assumption of Eq. (4), arbitrary observational summaries do not identify the additive backbone. For instance, after subtracting its own causal channel, Agent ρ and Agent σ have

$$R_\rho = X - f_\rho(Z_\rho) = f_\sigma(Z_\sigma) + \varepsilon, \qquad R_\sigma = X - f_\sigma(Z_\sigma) = f_\rho(Z_\rho) + \varepsilon.$$

These are agent-relative residuals, not common backbone residuals: R_ρ still bundles the other private response $f_\sigma(Z_\sigma)$ with the disturbance ε, while R_σ bundles $f_\rho(Z_\rho)$ with ε. A residual variance alone therefore does not separate variation due to the other causal channel from variation due to ε. What resolves the ambiguity is causal communication. Once Agent ρ communicates f_ρ and Agent σ communicates f_σ, the backbone mean response $m(z_\rho, z_\sigma) = f_\rho(z_\rho) + f_\sigma(z_\sigma)$ is fixed for every joint intervention pair, up to the shared constant $\mathbb{E}[\varepsilon]$ when absolute outcome levels rather than contrasts are required.

Example 3 (Education value-added under communication)**.** Following [5, 7], we specialise the additive response to a linear case and measure exposures in standard-deviation-year units: $f_\rho(z_\rho) = 1.3\, z_\rho$, $f_\sigma(z_\sigma) = 0.5\, z_\sigma$, with effects in percentage points. A full-childhood neighbourhood disadvantage of roughly one standard deviation, assuming a full childhood comprises eighteen one-year exposure periods and applying the per-year exposure estimate from [7] linearly over that exposure window, corresponds to about $\Delta_\sigma := 18 \times 0.5\% \approx 9\%$ lower adult earnings. If $\bar{z}_\rho$ is the baseline teacher-quality exposure and z_ρ^* is the compensating exposure, the offset solves

$$1.3\,(z_\rho^* - \bar{z}_\rho) = f_\rho(z_\rho^*) - f_\rho(\bar{z}_\rho) = \Delta_\sigma.$$

The required cumulative increase is therefore approximately $\Delta_\sigma / 1.3 \approx 9/1.3 \approx 7$ teacher-quality standard-deviation years. If spread evenly over a 13-year school career, this is about $7/13 \approx 0.54$ standard deviations above baseline per year. Within this linear additive special case, the communicated slopes suffice to identify the backbone response; if $Z_\rho \perp\!\!\!\perp Z_\sigma$, $\varepsilon \perp\!\!\!\perp (Z_\rho, Z_\sigma)$, and $\mathrm{Var}(X)$ is observed, the residual variance is recovered by $\mathrm{Var}(\varepsilon) = \mathrm{Var}(X) - \mathrm{Var}\big(f_\rho(Z_\rho)\big) - \mathrm{Var}\big(f_\sigma(Z_\sigma)\big)$. This calculation is not licensed by the local marginals alone. It becomes meaningful only after separability rules out the hidden interaction regime and the agents communicate causally identified response functions. Under those conditions, the neighbourhood study identifies the size of the disadvantage gap, the teacher study identifies the compensating instructional dose, and the RCK backbone becomes an operational object enabling causal inference from private reports.

3. Conclusion

RCK is a theory of causal transport given a shared *interventionally consistent* abstraction. This paper studies the identification problem that precedes such transport in private-knowledge common-effect settings. The main message is negative generically and positive conditionally. Private local causal marginals need not determine a unique edge-level backbone, and the ambiguity is causally meaningful because compatible backbones can disagree on joint interventions. Additive separability removes the interaction ambiguity, but agents must still communicate causal response objects rather than observational residual summaries.

This gives a narrower but actionable reading of RCK. The framework's transport maps are useful once the shared edge value is available. Our results clarify when private causal knowledge can supply that edge value, and when additional structure or communication is required. Natural next steps are to test separability in pooled studies, extend the result to larger agent networks, and identify weaker equivalence classes of backbones that preserve only the intervention queries needed for a given decision.

References

[1] J. Pearl, Causality, Cambridge University Press, 2009.

[2] J. Peters, D. Janzing, B. Schölkopf, Elements of Causal Inference: Foundations and Learning Algorithms, MIT Press, 2017.

[3] G. D'Acunto, C. Battiloro, The Relativity of Causal Knowledge, in: Proc. of UAI, 2025, pp. 863–881. URL: https://proceedings.mlr.press/v286/d-acunto25a.html.

[4] R. Chetty, J. N. Friedman, J. E. Rockoff, Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates, American Economic Review 104 (2014) 2593–2632. doi:10.1257/aer.104.9.2593.

[5] R. Chetty, J. N. Friedman, J. E. Rockoff, Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood, American Economic Review 104 (2014) 2633–2679. doi:10.1257/aer.104.9.2633.

[6] R. Chetty, N. Hendren, L. F. Katz, The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment, American Economic Review 106 (2016) 855–902. doi:10.1257/aer.20150572.

[7] R. Chetty, N. Hendren, The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates, The Quarterly Journal of Economics 133 (2018) 1163–1228. doi:10.1093/qje/qjy006.

[8] P. K. Rubenstein, S. Weichwald, S. Bongers, J. M. Mooij, D. Janzing, M. Grosse-Wentrup, B. Schölkopf, Causal consistency of structural equation models, in: Proc. of UAI, 2017, pp. 808–817.

[9] S. Beckers, J. Y. Halpern, Abstracting causal models, in: Proc. of AAAI, 2019, pp. 2678 – 2685.

[10] E. F. Rischel, S. Weichwald, Compositional abstraction error and a category of causal models, in: Proc. of UAI, 2021.

[11] A. Massidda, S. Beckers, E. Bareinboim, Causal abstraction with soft interventions, in: Proc. of UAI, 2023.

[12] G. D'Acunto, F. M. Zennaro, Y. Felekis, P. Di Lorenzo, Causal Abstraction Learning based on the Semantic Embedding Principle, in: Proc. of ICML, 2025.

[13] R. Bradley, F. Dietrich, C. List, Aggregating causal judgments, Philosophy of Science 81 (2014) 491–515. doi:10.1086/678044.

[14] D. Alrajeh, H. Chockler, J. Y. Halpern, Combining experts' causal judgments, in: Proc. of AAAI, 2018, pp. 6311–6318.

[15] M. Friedenberg, J. Y. Halpern, Combining the Causal Judgments of Experts with Possibly Different Focus Areas, in: Proc. of KR, 2018.

[16] J. Pearl, E. Bareinboim, External validity: From do-calculus to transportability across populations, Statistical Science 29 (2014) 579–595. doi:10.1214/14-STS486.

[17] E. Bareinboim, J. Pearl, Causal inference and the data-fusion problem, Proc. of the National Academy of Sciences 113 (2016) 7345–7352. doi:10.1073/pnas.1510507113.

[18] R. Xiong, A. Koenecke, M. Powell, Z. Shen, J. T. Vogelstein, S. Athey, Federated Causal Inference in Heterogeneous Observational Data, Statistics in Medicine 42 (2023) 4418–4439. doi:10.1002/sim.9868.

[19] L. Li, I. Ng, G. Luo, B. Huang, G. Chen, T. Liu, B. Gu, K. Zhang, Federated causal discovery from heterogeneous data, in: Proc. of ICLR, 2024. URL: https://openreview.net/forum?id=m7tJxajC3G.

[20] J. Tian, J. Pearl, On the Testable Implications of Causal Models with Hidden Variables, in: Proc. of UAI, 2002, pp. 519–527.

[21] L. Gresele, J. von Kügelgen, J. Kübler, M. Besserve, B. Schölkopf, Causal inference through the structural causal marginal problem, in: Proc. of ICML, 2022.

[22] S. P. Meyn, R. L. Tweedie, Markov Chains and Stochastic Stability, Springer-Verlag, London, 1993. URL: https://probability.ca/MT/. doi:10.1007/978-1-4471-3267-7.

[23] E. A. Hanushek, Conceptual and Empirical Issues in the Estimation of Educational Production Functions, Journal of Human Resources 14 (1979) 351–388. doi:10.2307/145575.

[24] J. E. Rockoff, The impact of individual teachers on student achievement: Evidence from panel data, American Economic Review 94 (2004) 247–252. doi:10.1257/0002828041302244.

[25] S. G. Rivkin, E. A. Hanushek, J. F. Kain, Teachers, schools, and academic achievement, Econometrica 73 (2005) 417–458. doi:10.1111/j.1468-0262.2005.00584.x.

[26] C. K. Jackson, Teacher Quality at the High-School Level: The Importance of Accounting for Tracks, Journal of Labor Economics 32 (2014) 645–684. doi:10.1086/676017.